The Annals of Statistics
1976, Vol. 4, No. 2, 292-301

APPROXIMATIONS FOR STATIONARY COVARIANCE MATRICES AND THEIR INVERSES WITH APPLICATION TO ARMA MODELS[1]

BY PAUL SHAMAN

Carnegie-Mellon University

Approximation of the covariance matrix $\mathbf{\Sigma}$ of T consecutive observations from a second-order stationary process with continuous positive spectral density $f(\lambda) = [\sigma^2/(2\pi)^2]|\sum_{j=0}^{\infty} \delta_j e^{i\lambda j}|^2$ is considered. If $\mathbf{\Sigma}^*$ is the covariance matrix corresponding to a process with spectral density $1/[(2\pi)^2 f(\lambda)]$, then $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{-1} \geqq 0$. A matrix $\sigma^{-2}\mathbf{A}'\mathbf{A}$ with the property that $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A} \geqq 0$ and $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1} \geqq 0$ is also considered. For autoregressive-moving average processes of order (p, q), $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ and $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$ are shown to have rank $\min[\max(p, q), T]$ and $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{-1}$ to have rank $\min[2\max(p, q), T]$. Some results concerning the covariance determinant are also discussed. If D_T is $\sigma^{-2T}|\mathbf{\Sigma}|$ for sample size T and $D_0 = 1$, then $D_T < D_{T+1}$, $T = 0, 1, \cdots$, unless the process is autoregressive of order p, in which case $1 < D_1 < \cdots < D_p = D_{p+1} = \cdots$.

1. Introduction. In time series studies approximation of the inverse of the covariance matrix of T consecutive observations from a stationary process is often of interest. This problem arises, for example, when an autoregressive-moving average (ARMA) process is assumed to be Gaussian and maximum likelihood estimation of its parameters is desired. Exact representations of the inverse are usually either unknown in closed form expressions or are too complicated to be useful when they are known. A list of references to known expressions for the exact inverse in various special cases of the ARMA model appears in Shaman (1975). See also Galbraith and Galbraith (1974) and Newbold (1974).

Consider a zero mean, second-order stationary stochastic process $\{x_t, t = 0, \pm 1, \cdots\}$ with continuous positive spectral density $f(\lambda)$. Let the covariance function be specified by

$$Ex_t x_{t+h} = \sigma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda)\, d\lambda\,, \qquad h = 0, \pm 1, \cdots. \tag{1.1}$$

Define

$$\sigma^*(h) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} e^{i\lambda h} f^{-1}(\lambda)\, d\lambda\,, \qquad h = 0, \pm 1, \cdots, \tag{1.2}$$

which is a covariance function. The corresponding spectral density is $1/[(2\pi)^2 f(\lambda)]$.

Let $\mathbf{\Sigma}$ denote the $T \times T$ covariance matrix of $\mathbf{x} = (x_1, \cdots, x_T)'$ and define $\mathbf{\Sigma}^*$

Received August 1974; revised June 1975.

[1] Research supported in part by the Office of Naval Research under Grant Number N00014-67-A-0314-0023.

AMS 1970 *subject classifications.* Primary 62M10; Secondary 60G10, 15A45.

Key words and phrases. Stationary process, covariance matrix, autoregressive-moving average (ARMA) process, positive semidefinite, rank of matrix, covariance determinant.

to be the $T \times T$ matrix with $\sigma^*(h)$ in place of $\sigma(h)$, $h = 0, 1, \cdots, T-1$. We shall study approximations to $\mathbf{\Sigma}^{-1}$. In particular, we shall consider a matrix $\sigma^{-2}\mathbf{A}'\mathbf{A}$ such that $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ and $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$ are positive semidefinite. Therefore $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{-1}$ is positive semidefinite (Theorem 2.2). These differences will be examined in some detail when $\{x_t\}$ is an ARMA process. In particular, their ranks will be studied. We shall also discuss approximation of $|\mathbf{\Sigma}|$.

ARMA models constitute a class of stationary processes which has recently been receiving considerable attention (see Box and Jenkins (1970) and Hannan (1970), e.g.). Let $\{\varepsilon_t\}$ be a sequence of uncorrelated random variables with mean 0 and variance σ^2. The process $\{x_t\}$ defined by

$$\sum_{j=0}^{p} \beta_j x_{t-j} = \sum_{k=0}^{q} \gamma_k \varepsilon_{t-k}, \qquad t = 0, \pm 1, \cdots, \tag{1.3}$$

with $\beta_0 = \gamma_0 = 1$, is an ARMA process of order (p, q). If $q = 0$ $\{x_t\}$ is also called an autoregressive process of order p, and if $p = 0$ it is called a moving average process of order q. We assume that $B(z) = 1 + \beta_1 z + \cdots + \beta_p z^p$ and $G(z) = 1 + \gamma_1 z + \cdots + \gamma_q z^q$ have no common zero and that for both all zeros are outside the unit circle. The spectral density of $\{x_t\}$ is $f(\lambda) = [\sigma^2/(2\pi)]|G(e^{i\lambda})|^2/|B(e^{i\lambda})|^2$.

In the general case where $f(\lambda)$ is positive and continuous $\{x_t\}$ admits the representation

$$x_t = \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j}, \qquad t = 0, \pm 1, \cdots, \tag{1.4}$$

with $\delta_0 = 1$, $\sum_{j=0}^{\infty} \delta_j^2 < \infty$. The spectral density of $\{x_t\}$ specified by (1.4) is $f(\lambda) = [\sigma^2/(2\pi)]|\sum_{j=0}^{\infty} \delta_j e^{i\lambda j}|^2$. In the case of (1.3) $\sum_{j=0}^{\infty} \delta_j z^j = B^{-1}(z)G(z)$ and the δ_j's are functions of $p + q$ parameters $\beta_1, \cdots, \beta_p, \gamma_1, \cdots, \gamma_q$. Sometimes it may be assumed that $\delta_j = \delta_j(\theta_1, \cdots, \theta_r)$, where the θ_i's are finitely many unknown parameters. The possibility that $r < p + q$ or that the θ_i's are not expressly the β_j's and γ_k's when (1.3) holds is not excluded.

When $\{x_t\}$ is Gaussian the density of $\mathbf{x}$ is

$$(2\pi)^{-\frac{1}{2}T}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x}\}. \tag{1.5}$$

Numerous authors have considered estimation of the parameters in (1.3) and/or the parameters characterizing the δ_j's in (1.4). The Gaussian assumption is not crucial for consistency if the estimation procedure merely originates from (1.5). However, some assumptions about fourth-order cumulants are needed to establish efficiency. Since $|\mathbf{\Sigma}|$ and $\mathbf{\Sigma}^{-1}$ are generally intractable and an explicit maximum likelihood solution cannot be found, the estimation studies have proceeded by maximizing an approximation to the likelihood (1.5). Whittle (1953, 1954) proposed replacing (1.5) by

$$\begin{aligned}(2\pi\sigma^2)^{-\frac{1}{2}T} \exp\left\{-\frac{T}{4\pi}\int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda)}\,d\lambda\right\} \\ = (2\pi)^{-\frac{1}{2}T} \exp\left\{-\frac{T}{4\pi}\int_{-\pi}^{\pi}\left[\log 2\pi f(\lambda) + \frac{I(\lambda)}{f(\lambda)}\right]d\lambda\right\},\end{aligned} \tag{1.6}$$

where

$$I(\lambda) = \frac{1}{2\pi T} |\sum_{t=1}^{T} x_t e^{i\lambda t}|^2$$

is the periodogram and we have used $2\pi \log \sigma^2 = \int_{-\pi}^{\pi} \log 2\pi f(\lambda)\, d\lambda$. The approximation (1.6) involves replacing $|\Sigma|$ by σ^{2T} and Σ^{-1} by Σ^*.

Estimation of parameters in ARMA models (1.3) has been treated using (1.6) or the same expression with a discrete sum in place of the integral by Whittle (1953, 1954), Durbin (1959), Hannan (1969, 1970), Clevenson (1970), Parzen (1971), and Anderson (1975c). Whittle (1962), Walker (1964), and Hannan (1973) have considered estimation of the parameters $\theta_1, \cdots, \theta_r$ characterizing the δ_j's in (1.4) by using (1.6). Box and Jenkins (1970) proceed in the case of (1.3) by maximizing by numerical means an approximation to the likelihood function which differs somewhat from (1.6) in the exponent of its first form. Anderson (1975a, b, c) has studied estimation of the parameters in (1.3) under the assumption $x_0 = \cdots = x_{-p+1} = \varepsilon_0 = \cdots = \varepsilon_{-q+1} = 0$. This leads to a modified likelihood function in which the covariance determinant is σ^{2T}, as it is for the modified likelihoods considered by all of the above authors. Mann and Wald (1943) treated $x_0, \cdots, x_{-p+1}$ as fixed values in considering parameter estimation for autoregressive processes.

2. Approximation of Σ^{-1}.

2.1. *Some general results.* The covariance function (1.1) of $\{x_t\}$ may be expressed as

$$\sigma(h) = \sigma^2 \sum_{j=0}^{\infty} \delta_j \delta_{j+|h|}, \qquad h = 0, \pm 1, \cdots.$$

We assume that the following expansion is valid for $|z| < 1 + \delta$, $\delta > 0$,

$$\sum_{j=0}^{\infty} \alpha_j z^j = (\sum_{j=0}^{\infty} \delta_j z^j)^{-1}. \tag{2.1}$$

Then (1.2) is also

$$\sigma^*(h) = \sigma^{-2} \sum_{j=0}^{\infty} \alpha_j \alpha_{j+|h|}, \qquad h = 0, \pm 1, \cdots,$$

with $\alpha_0 = 1$.

Let $\mathbf{L}$ be the $T \times T$ matrix with 1's on the diagonal directly below the main diagonal and 0's elsewhere and define

$$\mathbf{A} = \sum_{j=0}^{T-1} \alpha_j \mathbf{L}^j, \quad \mathbf{D} = \sum_{j=0}^{T-1} \delta_j \mathbf{L}^j.$$

Then $\mathbf{AD} = \mathbf{I}$. Let the following random variables be defined,

$$y_t = \sum_{j=0}^{\infty} \theta_{tj} \varepsilon_{t-j}, \qquad t = 1, \cdots, T,$$
$$\tilde{y}_t = \sum_{j=0}^{t-1} \theta_{tj} \varepsilon_{t-j}, \qquad t = 1, \cdots, T,$$

where the θ_{tj}'s are any coefficients such that the sums exist as limits in mean square. Denote the covariance matrix of $y_1, \cdots, y_T$ by $\mathbf{\Gamma}$ and that of $\tilde{y}_1, \cdots, \tilde{y}_T$ by $\tilde{\mathbf{\Gamma}}$.

THEOREM 2.1. $\mathbf{\Gamma} - \tilde{\mathbf{\Gamma}}$ *is positive semidefinite*, $T = 1, 2, \cdots$.

PROOF. $\Gamma - \tilde{\Gamma}$ is the covariance matrix of

$$\sum_{j=t}^{\infty} \theta_{tj} \varepsilon_{t-j}\,, \qquad t = 1, \cdots, T,\ T = 1, 2, \cdots .$$

COROLLARY 2.1. *$\Sigma - \sigma^2 \mathbf{D}\mathbf{D}'$ and $\Sigma^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ are positive semidefinite, $T = 1, 2, \cdots$.*

PROOF. The assertions follow directly from Theorem 2.1. Set $\theta_{tj} = \delta_j$ for the first matrix, and for the second set $\theta_{tj} = \alpha_j$.

Since Corollary 2.1 implies $\sigma^{-2}(\mathbf{D}')^{-1}\mathbf{D}^{-1} - \Sigma^{-1} = \sigma^{-2}\mathbf{A}'\mathbf{A} - \Sigma^{-1}$ is positive semidefinite, the following is an immediate consequence.

THEOREM 2.2. *Let Σ be the covariance matrix of T consecutive observations from a second-order stationary stochastic process with positive continuous spectral density $f(\lambda)$, and let Σ^* be the covariance matrix associated with a process with spectral density $1/[(2\pi)^2 f(\lambda)]$. Then $\Sigma^* - \Sigma^{-1}$ is positive semidefinite, $T = 1, 2, \cdots$.*

If $\Sigma^* - \Sigma^{-1}$ is positive definite for some value of T, then it is also positive definite for all smaller values.

THEOREM 2.3. *Let the conditions of Theorem 2.2 hold. If $\Sigma^* - \Sigma^{-1}$ is positive definite for $T = T_0$, then it is positive definite for $T = 1, \cdots, T_0 - 1$.*

PROOF. Partition Σ and Σ^{-1} according to

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix},$$

so that Σ_{11} and Σ^{11} are square and have the same dimension. Partition Σ^* similarly. Since $\Sigma_{11}^* - \Sigma^{11} > 0$ and $\Sigma_{11}^{-1} = \Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}$,

$$\begin{aligned} \Sigma_{11}^* - \Sigma_{11}^{-1} &= \Sigma_{11}^* - [\Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}] \\ &\geqq \Sigma_{11}^* - \Sigma^{11} > 0\,. \end{aligned}$$

For autoregressive processes of order p it is well known that Σ^{-1} is approximated by Σ^*. If $T \geqq 2p$ these two matrices are the same except in the $p \times p$ submatrices in the upper left and lower right corners. (See Section 6.2 of Anderson (1971), e.g.) This matter was examined in Shaman (1975), where it was shown that for $T > 2p$ columns (rows) $p + 1, \cdots, T - p$ of $\Sigma\Sigma^* - \mathbf{I}$ are zero vectors when $\{x_t\}$ is an autoregressive (a moving average) process of order p, and that the $2p$ nonzero characteristic roots of $\Sigma\Sigma^* - \mathbf{I}$ are positive. Thus Theorem 2.2 provides a partial generalization.

Corollary 2.1 and Theorem 2.2 imply that in the general case of $f(\lambda)$ positive and continuous $\sigma^{-2}\mathbf{A}'\mathbf{A}$ approximates Σ^{-1} better than Σ^* does (or at least as well). The expression used by Anderson (1975a, b, c) to approximate Σ^{-1} in arriving at a modified likelihood function for ARMA (p, q) processes is in fact $\sigma^{-2}\mathbf{A}'\mathbf{A}$. Below we shall have occasion to refer to

$$(2.2) \qquad \Sigma - \sigma^2\mathbf{D}\mathbf{D}' = \sigma^2\mathbf{E}\mathbf{E}' = [\sigma^2 \textstyle\sum_{j=\min(s,t)}^{\infty} \delta_j \delta_{j+|s-t|}]_{s,t=1,\cdots,T}\,,$$

where the element in row t and column j of $\mathbf{E}$ is δ_{t+j-1}, $t = 1, \cdots, T$, $j = 1, 2, \cdots$, and also

$$(2.3) \qquad \mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A} = \sigma^{-2}\mathbf{F}\mathbf{F}' = [\sigma^{-2} \textstyle\sum_{j=\min(T-s+1,T-t+1)}^{\infty} \alpha_j \alpha_{j+|s-t|}]_{s,t=1,\cdots,T} ,$$

where the element in row t and column j of $\mathbf{F}$ is α_{T-t+j}, $t = 1, \cdots, T$, $j = 1, 2, \cdots$.

Consider again as an example the autoregressive process of order p. Then $\alpha_j = \beta_j$, $j = 0, 1, \cdots, p$, $\alpha_j = 0$, $j = p + 1, p + 2, \cdots$, and $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ has a $p \times p$ block in the lower right corner and 0's elsewhere if $T \geqq p$. If $T \geqq 2p$ it follows immediately from the known form for $\mathbf{\Sigma}^{-1}$ (see Siddiqui (1958), e.g.) that $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$ is the transpose about the transverse diagonal of $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$. This relation also holds for $p \leqq T < 2p$, as may be deduced from examination of

$$(2.4) \qquad \textstyle\sum_{s,t=1}^{T} \sigma_T^{st} x_s x_t = \sum_{s,t=1}^{p} \sigma_p^{st} x_s x_t + \sigma^{-2} \sum_{t=p+1}^{T} (x_t + \beta_1 x_{t-1} + \cdots + \beta_p x_{t-p})^2 ,$$

where $\mathbf{\Sigma}^{-1} = (\sigma_T^{st})$, and use of (see Siddiqui (1958))

$$\sigma_p^{st} = \sigma_p^{p+1-t,p+1-s} = \textstyle\sum_{j=0}^{\min(s,t)-1} [\beta_j \beta_{j+|s-t|} - \beta_{p-j} \beta_{p-j-|s-t|}] , \qquad 2 \leqq s + t \leqq p + 1 .$$

The second summation on the right-hand side of (2.4) is not present if $T = p$. If $T < p$ $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ and $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$ are not transposes about the transverse diagonal of each other. These same relations hold for moving average processes of finite order when applied to $\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$ and $\sigma^2\mathbf{D}\mathbf{D}' - \mathbf{\Sigma}^{*-1}$, except that the lower right corner is replaced by the upper left corner. For ARMA (p, q) processes with $p > 0$ and $q > 0$ only the rank conditions described below hold generally.

2.2. *Autoregressive-moving average processes.* We examine $\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$ and $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ in more detail when $\{x_t\}$ is an ARMA (p, q) process as defined at (1.3). For such a process

$$(2.5) \qquad \textstyle\sum_{j=0}^{\min(p,r)} \beta_j \delta_{r-j} = 0 , \qquad r = q + 1, q + 2, \cdots .$$

With the aid of (2.5) Corollary 2.1 can be refined.

THEOREM 2.4. *Let* $\{x_t\}$ *be an* ARMA (p, q) *process with* $\beta_p \neq 0$ *and* $\gamma_q \neq 0$. *Then* $\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$ *and* $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ *have rank* $\min[\max(p, q), T]$.

PROOF. We present the details of the proof for $\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$.

(a) $T \geqq \max(p, q)$. First assume $p \geqq q$. Then if $T > p$ row r of $\mathbf{E}$ defined in (2.2) is a linear combination of rows $r - 1, \cdots, r - p$, $r = p + 1, \cdots, T$. This follows directly from (2.5) for $p \geqq q$. Moreover, the first p rows of $\mathbf{E}$ are linearly independent. For if they are not there exist $b_0, \cdots, b_{p-1}$, not all 0, such that

$$(2.6) \qquad \textstyle\sum_{j=0}^{p-1} b_j \delta_{r-j} = 0 , \qquad r = p, p + 1, \cdots .$$

The left-hand side has transform

$$\textstyle\sum_{j=0}^{p-1} b_j z^j \sum_{k=0}^{\infty} \delta_k z^k = \sum_{j=0}^{p-1} b_j z^j (\sum_{k=0}^{p} \beta_k z^k)^{-1} \sum_{l=0}^{q} \gamma_l z^l ,$$

which is a polynomial of degree at most $p-1$. This is impossible unless $b_0 = \cdots = b_{p-1} = 0$, because no zero of $B(z)$ is a zero of $G(z)$. Now suppose $q > p$. In this case if $T > q$ row r of $\mathbf{E}$ is a linear combination of rows $r-1, \cdots, r-p$, $r = q+1, \cdots, T$. Rows $1, \cdots, q-p$ are not involved—they relate to end effects. This assertion follows directly from (2.5). Also, the first q rows of $\mathbf{E}$ are linearly independent. If they are not, there exist $b_0, \cdots, b_{q-1}$, not all 0, such that (2.6) holds with p replaced by q. The left-hand side of this modified (2.6) has transform

$$\sum_{j=0}^{q-1} b_j z^j \left(\sum_{k=0}^{p} \beta_k z^k\right)^{-1} \sum_{l=0}^{q} \gamma_l z^l , \tag{2.7}$$

which is a polynomial of degree at most $q-1$. Since $q > p$, select $b_0, \cdots, b_{q-1}$ so that $B(z)$ divides $\sum_{j=0}^{q-1} b_j z^j$. Then let (2.7) be

$$\sum_{j=0}^{q-p-1} c_j z^j \sum_{l=0}^{q} \gamma_l z^l , \tag{2.8}$$

still a polynomial of degree at most $q-1$. Thus

$$\begin{aligned} 0 &= c_{q-p-1}\gamma_q , \\ 0 &= c_{q-p-1}\gamma_{q-1} + c_{q-p-2}\gamma_q , \\ &\vdots \\ 0 &= c_{q-p-1}\gamma_{p+1} + \cdots + c_0\gamma_q . \end{aligned} \tag{2.9}$$

Since $\gamma_q \neq 0$, we must have $c_0 = \cdots = c_{q-p-1} = 0$, in which case $b_0 = \cdots = b_{q-1} = 0$.

(b) $T < \max(p, q)$. By (a) $\mathbf{\Sigma} - \sigma^2\mathbf{DD}' = \sigma^2\mathbf{EE}'$ is positive definite if $T = \max(p, q)$, and every principal minor is therefore also positive definite.

That the rank of $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ is also $\min[\max(p, q), T]$ follows in a similar manner by considering (2.3). The analog of (2.5) is

$$\sum_{j=0}^{\min(q,r)} \gamma_j \alpha_{r-j} = 0 , \qquad r = p+1, p+2, \cdots .$$

For autoregressive and moving average processes the rank result of Theorem 2.4 is an immediate consequence of the lemma in Shaman (1975) when T is at least twice the order.

Let us consider an example. The process $x_t + \beta x_{t-1} = \varepsilon_t + \gamma\varepsilon_{t-1}$ has $\delta_0 = \alpha_0 = 1$, $\delta_j = (\gamma - \beta)(-\beta)^{j-1}$, $\alpha_j = (\beta - \gamma)(-\gamma)^{j-1}$, $j = 1, 2, \cdots$. Then

$$\mathbf{\Sigma} - \sigma^2\mathbf{DD}' = \frac{\sigma^2(\gamma - \beta)^2}{1 - \beta^2} \begin{bmatrix} 1 \\ -\beta \\ \vdots \\ (-\beta)^{T-1} \end{bmatrix} [1 \; -\beta \cdots (-\beta)^{T-1}]$$

and

$$\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A} = \frac{(\beta - \gamma)^2}{\sigma^2(1 - \gamma^2)} \begin{bmatrix} (-\gamma)^{T-1} \\ (-\gamma)^{T-2} \\ \vdots \\ 1 \end{bmatrix} [(-\gamma)^{T-1}(-\gamma)^{T-2} \cdots 1] .$$

If $\{x_t\}$ is an autoregressive process of order p and $T \geqq 2p$ then $\boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}^{-1}$ has rank $2p$. This follows from the discussion directly below Theorem 2.3 and the result may be generalized.

THEOREM 2.5. *Under the conditions of Theorem* 2.4 *and* $T \geqq 2 \max(p, q)$, *no column* (*row*) *of* $\boldsymbol{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$ *is a linear combination of the columns* (*rows*) *of* $\boldsymbol{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$.

PROOF. We can seek constraints on $\mathbf{y} = (y_0, \cdots, y_{T-1})'$ such that $\mathbf{y}'(\sigma^{-2}\mathbf{F}\mathbf{F}' + \sigma^2\mathbf{E}\mathbf{E}')\mathbf{y} = 0$ [see (2.2) and (2.3)]. Let $q \geqq p$. First consider $\mathbf{y}'\mathbf{F} = \mathbf{0}$, or

$$\sum_{j=0}^{T-1} y_j \alpha_{r-j} = 0 , \qquad r = T, T+1, \cdots .$$

The transform of the left-hand side is

$$\sum_{j=0}^{T-1} y_j z^j \sum_{k=0}^{p} \beta_k z^k \left(\sum_{l=0}^{q} \gamma_l z^l\right)^{-1} , \tag{2.10}$$

a polynomial of degree at most $T-1$. Therefore $\mathbf{y}$ must be chosen so that each zero of $G(z)$ is a zero of $Y(z) = \sum_{j=0}^{T-1} y_j z^j$. Then (2.10) is

$$\sum_{j=0}^{T-1-q} c_j z^j \sum_{k=0}^{p} \beta_k z^k , \tag{2.11}$$

which is certainly a polynomial of degree at most $T-1$. Next consider $\mathbf{y}'\mathbf{E} = \mathbf{0}$, or

$$\sum_{j=0}^{T-1} y_{T-1-j} \delta_{r-j} = 0 , \qquad r = T, T+1, \cdots .$$

The transform is

$$\sum_{j=0}^{T-1} y_{T-1-j} z^j \left(\sum_{k=0}^{p} \beta_k z^k\right)^{-1} \sum_{l=0}^{q} \gamma_l z^l , \tag{2.12}$$

a polynomial of degree at most $T-1$, and $\mathbf{y}$ must be chosen so that each zero of $B(z)$ is a zero of $\sum_{j=0}^{T-1} y_{T-1-j} z^j$. In that case (2.12) is

$$\sum_{j=0}^{T-1-p} d_j z^j \sum_{l=0}^{q} \gamma_l z^l . \tag{2.13}$$

In order that this be a polynomial of degree at most $T-1$, $d_{T-q} = \cdots = d_{T-1-p} = 0$ is necessary (if in fact $q > p$), because $\gamma_q \neq 0$. Then (2.10)—(2.13) imply

$$Y(z) = G(z) \sum_{j=0}^{T-1-q} c_j z^j , \tag{2.14}$$

$$z^{T-1} Y(z^{-1}) = B(z) \sum_{j=0}^{T-1-q} d_j z^j . \tag{2.15}$$

But (2.15) implies $y_0 = \cdots = y_{q-p-1} = 0$ (if $q > p$) and therefore by (2.14) $c_0 = \cdots = c_{q-p-1} = 0$. Thus each zero of $G(z)$ is a zero of $Y(z)$, the reciprocal of each zero of $B(z)$ is a zero of $Y(z)$, and $Y(z)$ has $q - p$ zeros with the value 0. Therefore $T - q - p - (q - p) = T - 2q$ coordinates of $\mathbf{y}$ can be chosen freely. It follows that $\boldsymbol{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A} + \boldsymbol{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$ has rank $2q$.

A similar proof holds if $p > q$.

COROLLARY 2.2. *Under the conditions of Theorem* 2.4 $\boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}^{-1}$ *has rank* $\min[2 \max(p, q), T]$.

PROOF. Consider $T \geqq 2 \max(p, q)$. Since

$$\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1} = \sigma^{-2}\mathbf{\Sigma}^{-1}(\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}')\mathbf{A}'\mathbf{A} ,$$

the theorem follows from Theorem 2.5 for $T \geqq 2 \max(p, q)$. Then apply Theorem 2.3 to cover $T < 2 \max(p, q)$.

Finally, the next theorem ties together Corollary 2.1, Theorem 2.2, and Theorem 2.3 on the one hand and Theorem 2.4 and Corollary 2.2 on the other.

THEOREM 2.6. *Let the conditions of Theorem* 2.2 *hold. If* $\{x_t\}$ *is not an* ARMA (p, q) *process with* p *and* q *both finite, then* $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$, $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$, *and* $\mathbf{\Sigma}^* - \mathbf{\Sigma}^{-1}$ *are positive definite for* $T = 1, 2, \cdots$.

PROOF. Consider $\mathbf{\Sigma}^* - \sigma^{-2}\mathbf{A}'\mathbf{A}$. The rows of $\mathbf{F}$ [(2.3)] are linearly independent for all T. If they are not, there would exist $b_0, \cdots, b_{T-1}$, not all 0, such that

$$\sum_{j=0}^{T-1} b_j \alpha_{r-j} = 0 , \qquad r = T, T+1, \cdots .$$

But

$$\sum_{j=0}^{T-1} b_j z^j \sum_{k=0}^{\infty} \alpha_k z^k$$

can be a polynomial of degree at most $T - 1$ only if $b_0 = \cdots = b_{T-1} = 0$. Similarly, $\mathbf{\Sigma} - \sigma^2\mathbf{D}\mathbf{D}'$, and therefore $\sigma^{-2}\mathbf{A}'\mathbf{A} - \mathbf{\Sigma}^{-1}$, are positive definite for all T.

3. The covariance determinant. In this section we discuss briefly approximation of the covariance determinant. Some of the results stated are known. As (1.6) indicates, it is customary to approximate $|\mathbf{\Sigma}|$ by σ^{2T}.

Let $\mathbf{\Sigma}_T$ denote the $T \times T$ covariance matrix of $(x_1, \cdots, x_T)'$ formed from (1.1) and assume (2.1) is valid for $|z| < 1 + \delta$, $\delta > 0$. Let $\mathbf{\Sigma}_T^*$ denote the corresponding matrix formed from (1.2). The representation (1.4) holds with $\operatorname{Var} \varepsilon_t = \sigma^2$.

Corollary 2.1 implies $|\mathbf{\Sigma}_T^*|^{-1} \leqq \sigma^{2T} \leqq |\mathbf{\Sigma}_T|$, $T = 1, 2, \cdots$. Theorem 2.4 implies that these inequalities are both strict if $\{x_t\}$ is an ARMA (p, q) process with $\beta_p \neq 0$ and $\gamma_q \neq 0$ and $T \leqq \max(p, q)$. Moreover, by Theorem 2.6 the inequalities are strict for all T if $\{x_t\}$ is not ARMA (p, q) with p and q both finite. A more precise result is given by Theorem 3.1 below.

Let $D_T = \sigma^{-2T}|\mathbf{\Sigma}_T|$, $T = 1, 2, \cdots$, and $D_0 = 1$. Then $D_T \leqq D_{T+1}$, $T = 0, 1, \cdots$. See Grenander and Szegö (1958), page 76. This result also holds for $D_T^* = \sigma^{2T}|\mathbf{\Sigma}_T^*|$. In fact strict inequality prevails unless $\{x_t\}$ is autoregressive of finite order.

THEOREM 3.1. *If* $\{x_t\}$ *is autoregressive of order* $p(\beta_p \neq 0)$, *then* $D_T < D_{T+1}$, $T = 0, 1, \cdots, p-1$, *and* $D_p = D_{p+1} = \cdots$. *If* $\{x_t\}$ *is not autoregressive of finite order, then* $D_T < D_{T+1}$, $T = 0, 1, \cdots$.

PROOF. Partition $\mathbf{\Sigma}_{k+1}$ as

$$\mathbf{\Sigma}_{k+1} = \begin{bmatrix} \sigma(0) & \boldsymbol{\sigma}_k' \\ \boldsymbol{\sigma}_k & \mathbf{\Sigma}_k \end{bmatrix}, \qquad k = 1, 2, \cdots .$$

Let $\{x_t\}$ be autoregressive of order p and write

$$D_{p+j} = \frac{\prod_{k=p}^{p+j-1} [\sigma(0) - \boldsymbol{\sigma}_k' \Sigma_k^{-1} \boldsymbol{\sigma}_k]}{\sigma^{2j}} D_p , \qquad j = 1, 2, \cdots .$$

The residual variance of x_t on $x_{t-1}, \cdots, x_{t-p}$ is equal to the residual variance of x_t on $x_{t-1}, \cdots, x_{t-k}$, $k = p+1, p+2, \cdots$. Thus $D_T < D_{T+1}$, $T = 0, 1, \cdots$, if $\{x_t\}$ is not autoregressive of finite order. Grenander and Szegö (1958) give $D_p = D_{p+1} = \cdots$ on page 71.

If $\{x_t\}$ is autoregressive of order p or a moving average of order q, Finch (1960) and Walker (1961) have shown that $\lim_{T\to\infty} D_T = \lim_{T\to\infty} D_T^*$. Grenander and Szegö (1958), page 76, evaluate the limit. Hannan (1973) proves $D_T \geqq 1$, $T = 0, 1, \cdots$, and

$$\lim_{T\to\infty} (1/T) \log D_T = 0 .$$

The discussion in this section has focused on results which are valid for every sample size T.

Acknowledgment. The author thanks T. W. Anderson for helpful comments and for suggestions which improved the organization of this paper.

REFERENCES

[1] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

[2] ANDERSON, T. W. (1975a). On maximum likelihood estimation of parameters of autoregressive moving average processes. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Theory, and Random Processes.*

[3] ANDERSON, T. W. (1975b). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. *Ann. Statist.* **3** 1283–1304.

[4] ANDERSON, T. W. (1975c). Estimation by maximum likelihood in autoregressive moving average models in the time and frequency domains. Technical Report, Department of Statistics, Stanford Univ.

[5] BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control.* Holden-Day, San Francisco.

[6] CLEVENSON, M. L. (1970). Asymptotically efficient estimates of the parameters of a moving average time series. Technical Report No. 15, Dept. of Statist., Stanford Univ.

[7] DURBIN, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46** 306–316.

[8] FINCH, P. D. (1960). On the covariance determinants of moving average and autoregressive models. *Biometrika* **47** 194–196.

[9] GALBRAITH, R. F. and GALBRAITH, J. I. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *J. Appl. Probability* **11** 63–71.

[10] GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications.* Univ. of California Press.

[11] Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems. *Biometrika* **56** 579–593.

[12] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

[13] HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probability* **10** 130–145.

[14] MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.

[15] NEWBOLD, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika* **61** 423–426.

[16] PARZEN, E. (1971). Efficient estimation of stationary time series mixed schemes. *Bull. Inst. Internat. Statist.* **44** 315–319.

[17] SHAMAN, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Ann. Statist.* **3** 532–538.

[18] SIDDIQUI, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.* **29** 585–588.

[19] WALKER, A. M. (1961). On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving average process. *Biometrika* **48** 197–199.

[20] WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary nondeterministic time-series. *J. Austral. Math. Soc.* **4** 363–384.

[21] WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

[22] WHITTLE, P. (1954). Appendix 2 to *A Study in the Analysis of Stationary Time Series* by H. Wold. Almqvist and Wiksell, Uppsala.

[23] WHITTLE, P. (1962). Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.* **39** 105–129.

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA 15213